09/703,574

OIPE JC139
JUL 20 2005
PATENT & TRADEMARK OFFICE

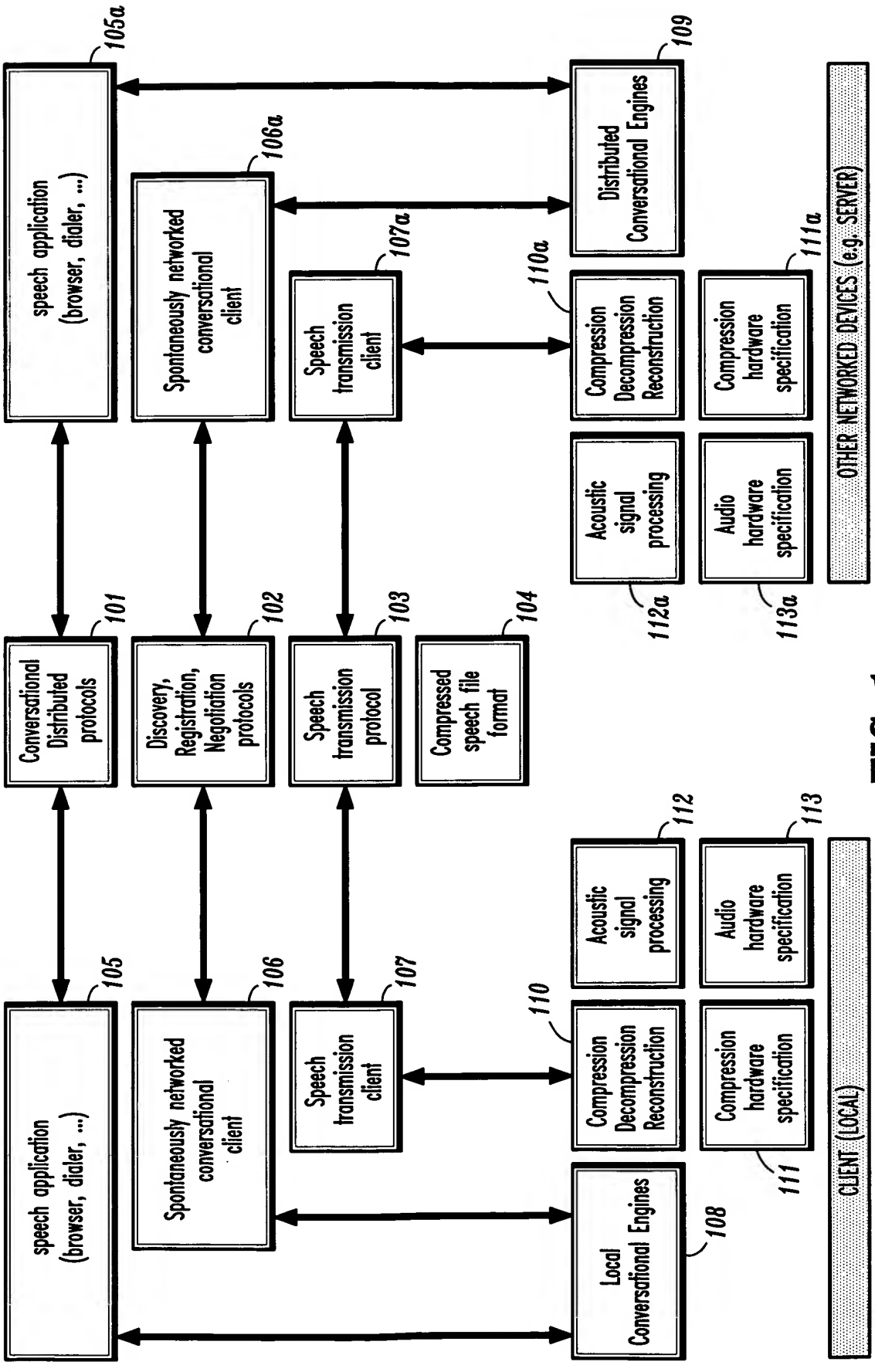

FIG. 1

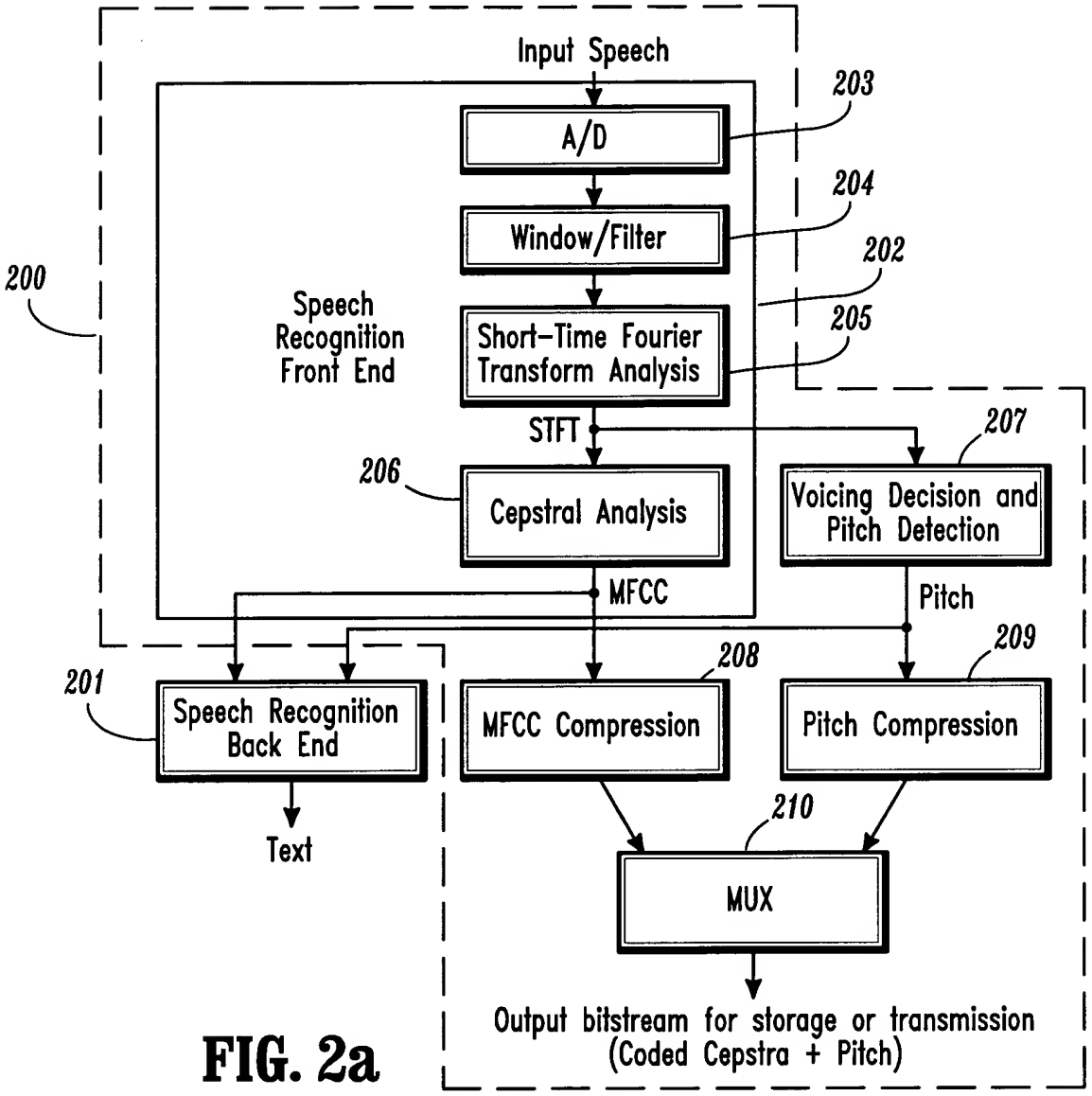
Input Speech
203
A/D
204
Window/Filter
202
200
Speech Recognition Front End
205
Short-Time Fourier Transform Analysis
STFT
207
206
Cepstral Analysis
Voicing Decision and Pitch Detection
MFCC
Pitch
209
208
201
Speech Recognition Back End
MFCC Compression
Pitch Compression
Text
210
MUX
Output bitstream for storage or transmission
(Coded Cepstra + Pitch)

FIG. 2a

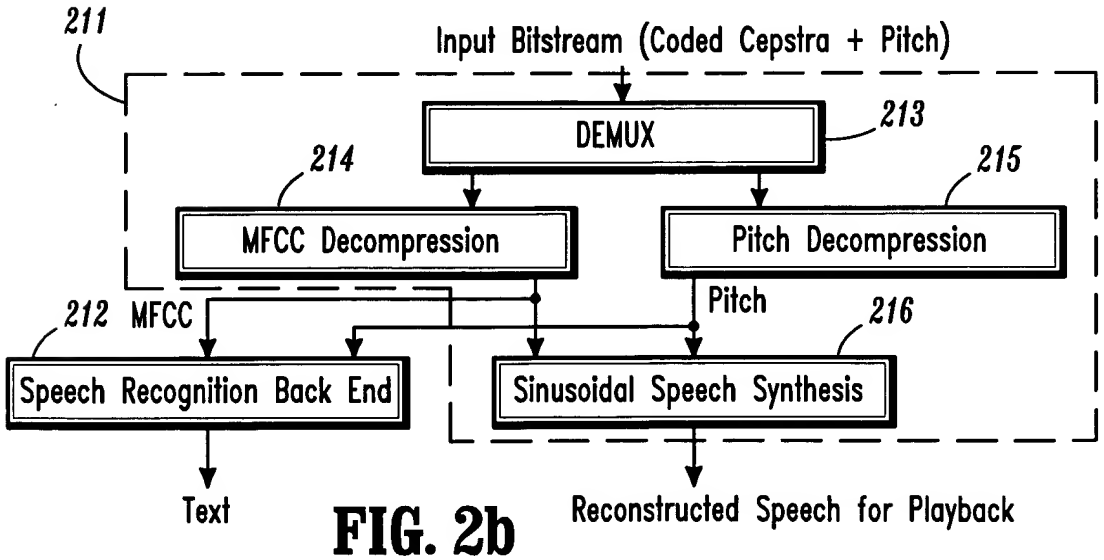
211
Input Bitstream (Coded Cepstra + Pitch)
DEMUX
213
214
215
MFCC Decompression
Pitch Decompression
Pitch
212 MFCC
216
Speech Recognition Back End
Sinusoidal Speech Synthesis
Text
Reconstructed Speech for Playback

FIG. 2b

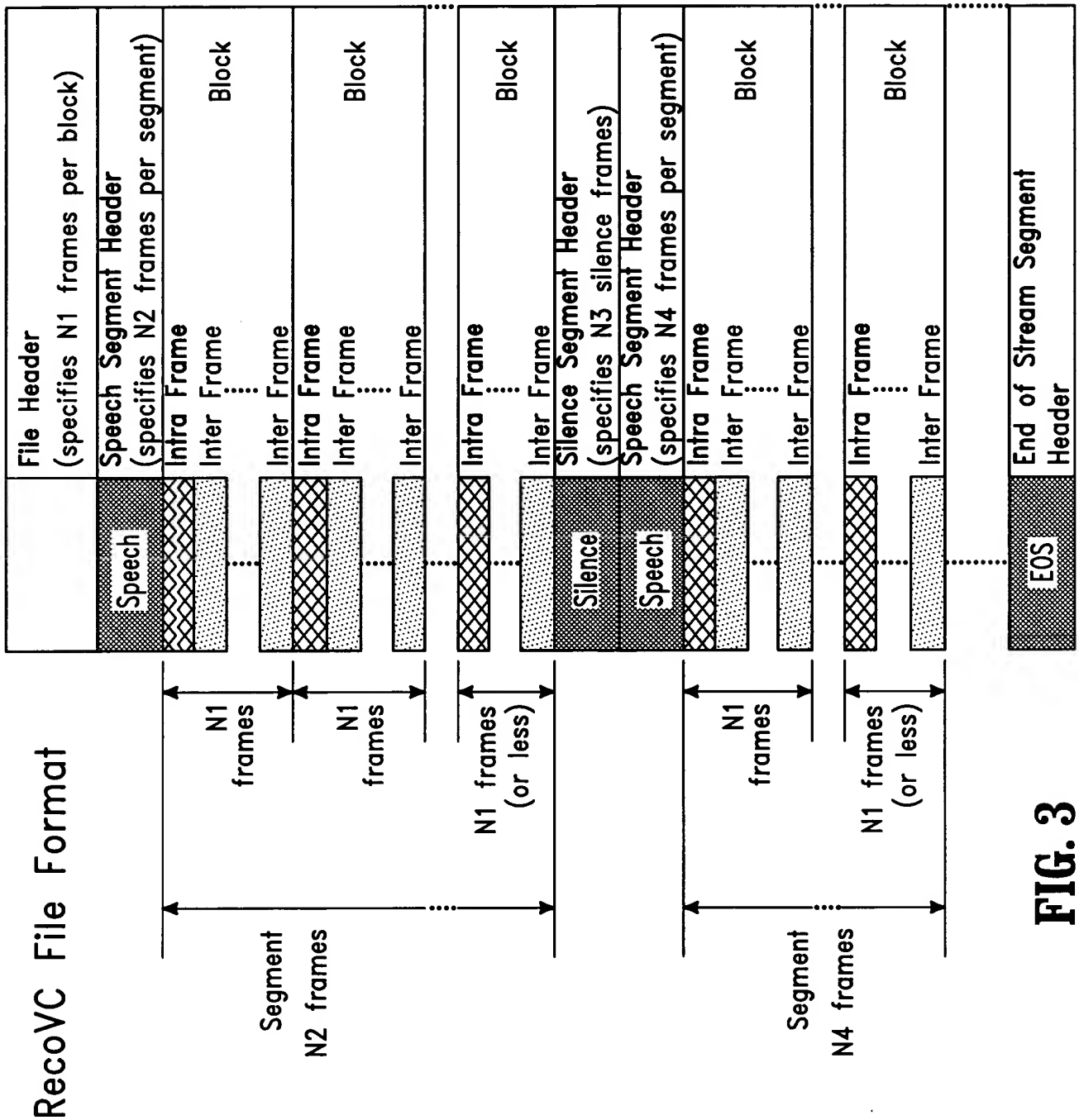
RecoVC File Format
File Header
(specifies N1 frames per block)
Speech Segment Header
(specifies N2 frames per segment)
Intra Frame
Inter Frame
Block
Speech
Silence Segment Header
(specifies N3 silence frames)
Silence
Speech Segment Header
(specifies N4 frames per segment)
End of Stream Segment Header
EOS
N1 frames
N1 frames (or less)
Segment N2 frames
Segment N4 frames

FIG. 3

FIG. 4–RecoVC File Header

Field Name	M/O	Number of Bytes	Value(s)	Description
Recovc magic number	M	4 bytes	0x2E727663	A fixed value which equals ".rvc" in ASCII representation. Used to verify that this is indeed a Recovc file.
Header length	M	2 byte	13–(2^16–1)	Total number of bytes in this header
Version Major	M	1 byte	0–255	File format major version number
Version Minor	M	1 byte	0–255	File format minor version number
Frame duration	M	1 byte	0x00	Index into the frame duration table below.
Frames per block	M	1 byte	1–255	Each block can contain one intra–frame only, or one intra–frame with up to 244 additional inter–frames.
Sampling rate	M	1 byte	0x00, 0x01 or 0x02	Index into sampling rate table below.
Cepstra dimension	M	1 byte	0x00 or 0x01	Index into Cepstra vector dimension I type below
Language	M	1 byte	0x00	Index into language table below
Profile	M	1 byte	0x00 or 0x01	Index into profile table below
Total frames	O	4 bytes	0–(2^32–2) or 0xFFFFFFFF	Total number of audio frames in this file, INCLUDING silence frames.
Total speech frames	O	4 bytes	0–(2^32–2) or 0xFFFFFFFF	Total number of audio frames in this file, EXCLUDING silence frames.
<Ancillary Data>	O			Reserved for future usage

09/703,576

Frame Duration Table

Code	Duration (msec)
0x00	10
0x01–0xFF	Reserved

Sampling Rate Table

Code	Input Sampling Rate [Hz]
0x00	8000
0x01	11025 (or 11000)
0x02	16000
0x03	22050 (or 22000)
0x04–0xFF	Reserved

Cepstra Vector Dimension/type Table

Code	Value
0x00	13
0x01	24
0x02–0xFF	Reserved

Language Table

Code	Description
0x00	English
0x01–0xFF	Reserved

FIG. 5

Profile Table

Code	Description
0x00	Recognition only–Only Cepstra is available in this file
0x01	Reconstruction enabled–Cepstra and pitch is available in the file
0x02–0xFF	Reserved

Speech and Silence Segment Headers (5 Byte)

Field Name	M/O	Number of bytes	Value(s)	Description
Segment type	M	1 byte	0x00 or 0x01	Index into segment type table below (speech or silence)
Number of frames	M	4 bytes	0–(2^31–2) or 0xFFFFFFFF	The total number of frames in this segment.

End of Stream and Data Segment Headers

Field Name	M/O	Number of bytes	Value(s)	Description
Segment type	M	1 byte	0x02 or 0x03–0xFF	Index into segment type table below (end of stream or data)
Segment length	M	4 bytes	5–(2^31–1)	The total number of bytes in this segment.
<Ancillary Data>	O			Place optional data here

FIG. 6

09/703,574

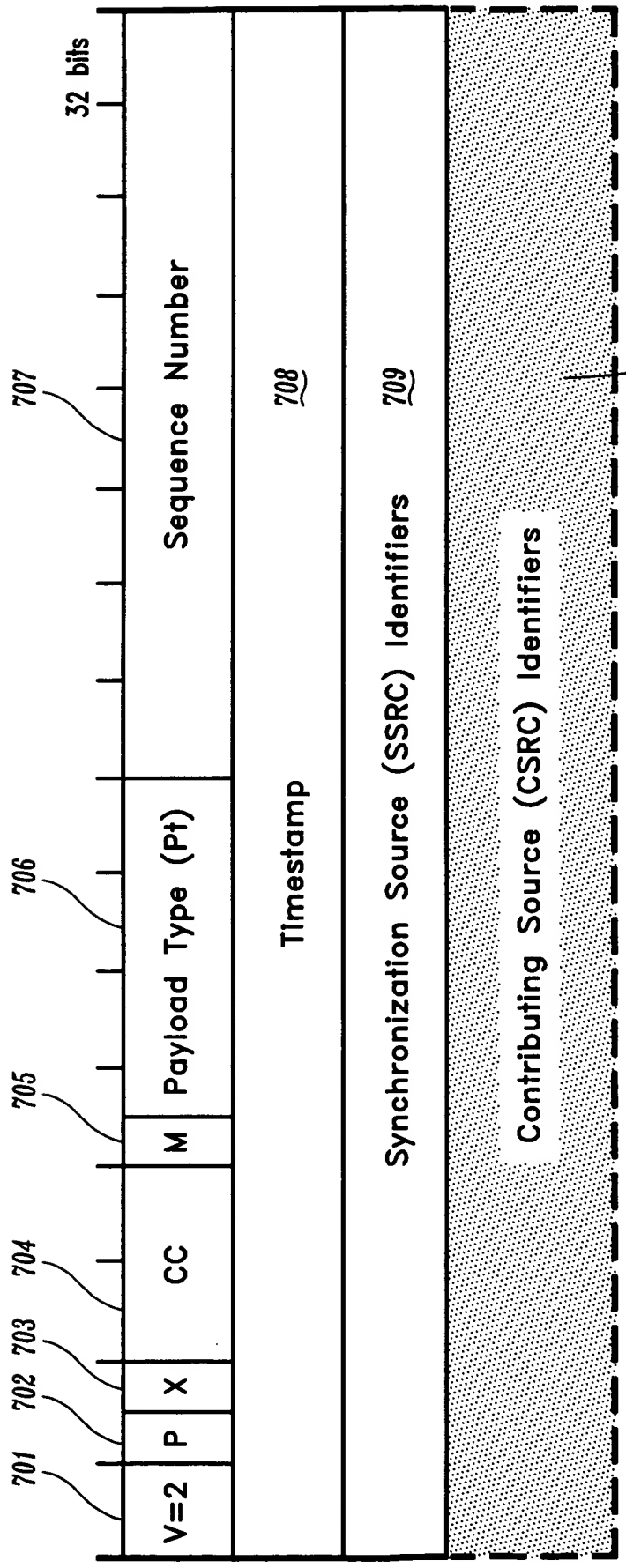

FIG. 7 (Prior Art)

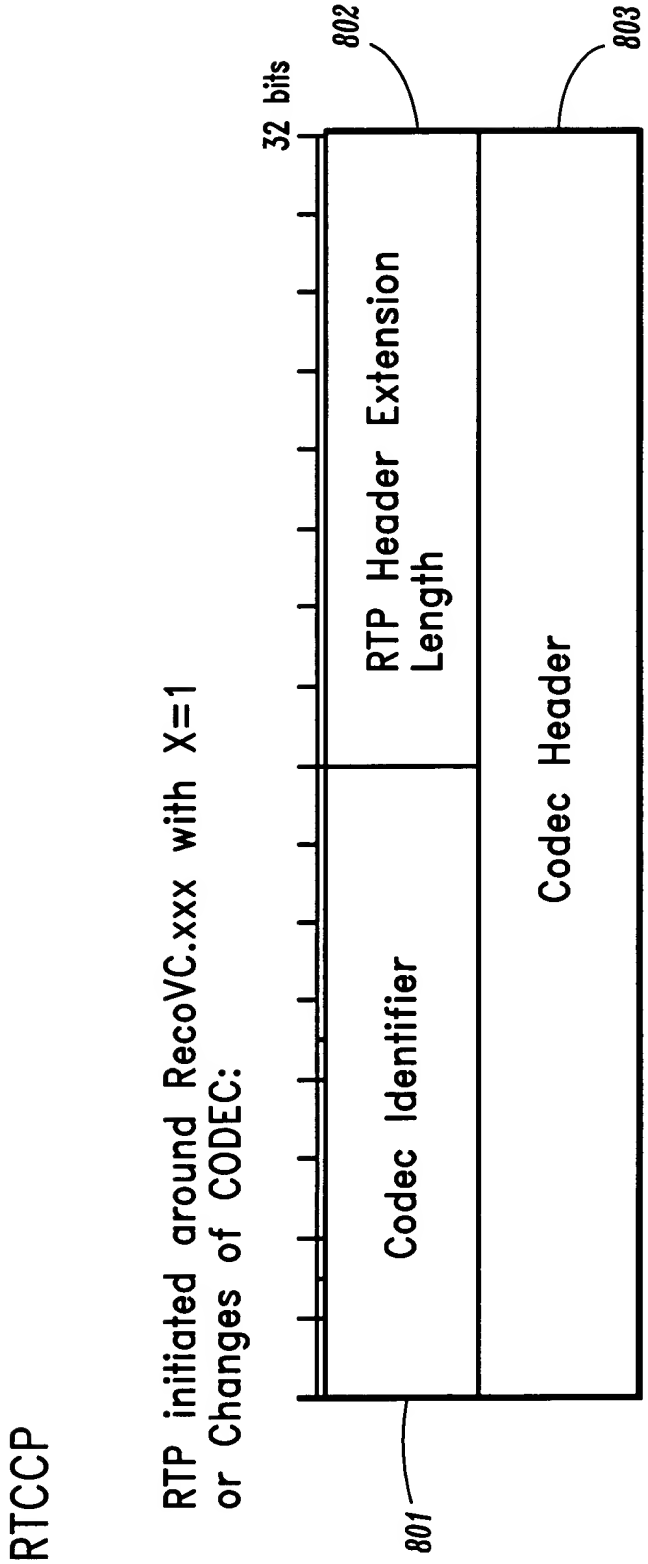
RTCCP
RTP initiated around RecoVC.xxx with X=1
or Changes of CODEC:
32 bits
802
803
RTP Header Extension Length
Codec Identifier
Codec Header
801
800

FIG. 8

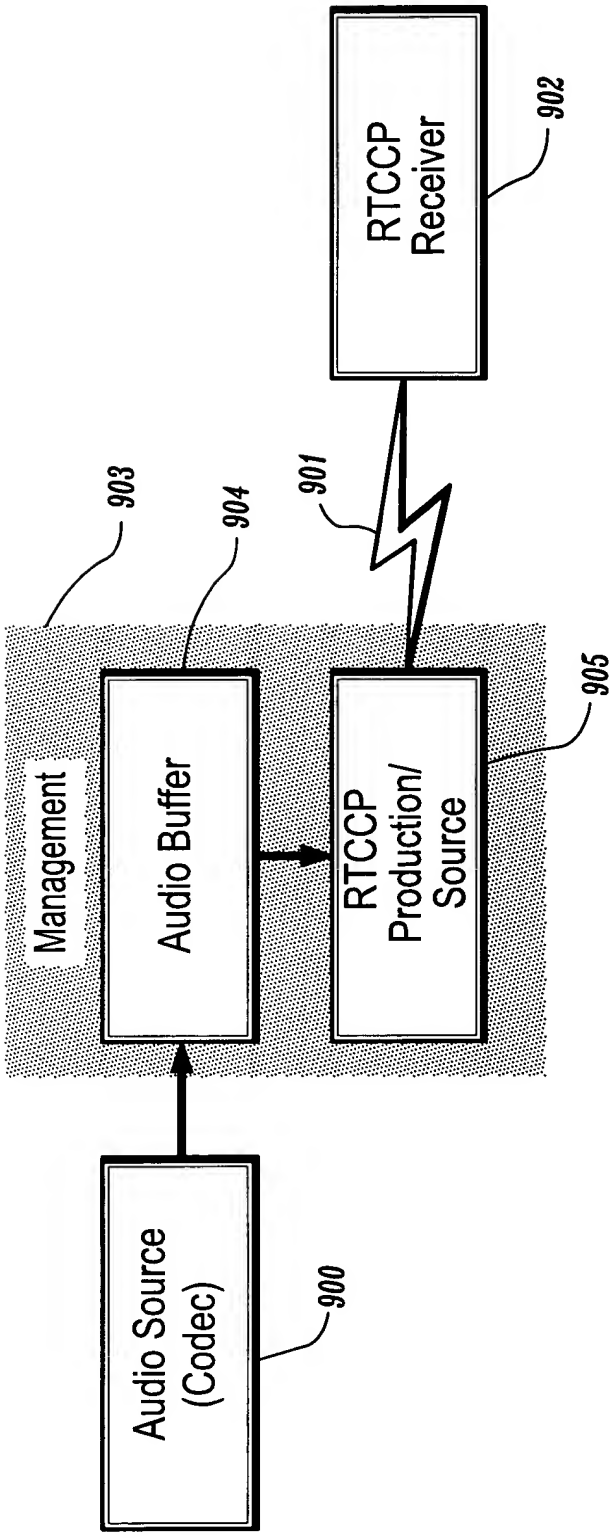
RTCCP Receiver
902
901
903
904
Management
Audio Buffer
RTCCP Production/ Source
905
Audio Source (Codec)
900

FIG. 9

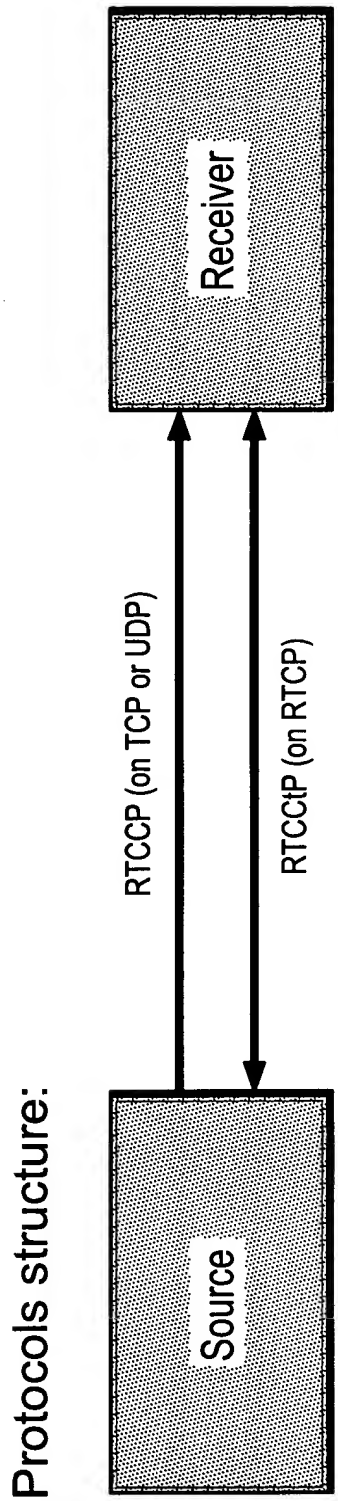
Protocols structure:
Receiver
RTCCP (on TCP or UDP)
RTCCtP (on RTCP)
Source

FIG. 11

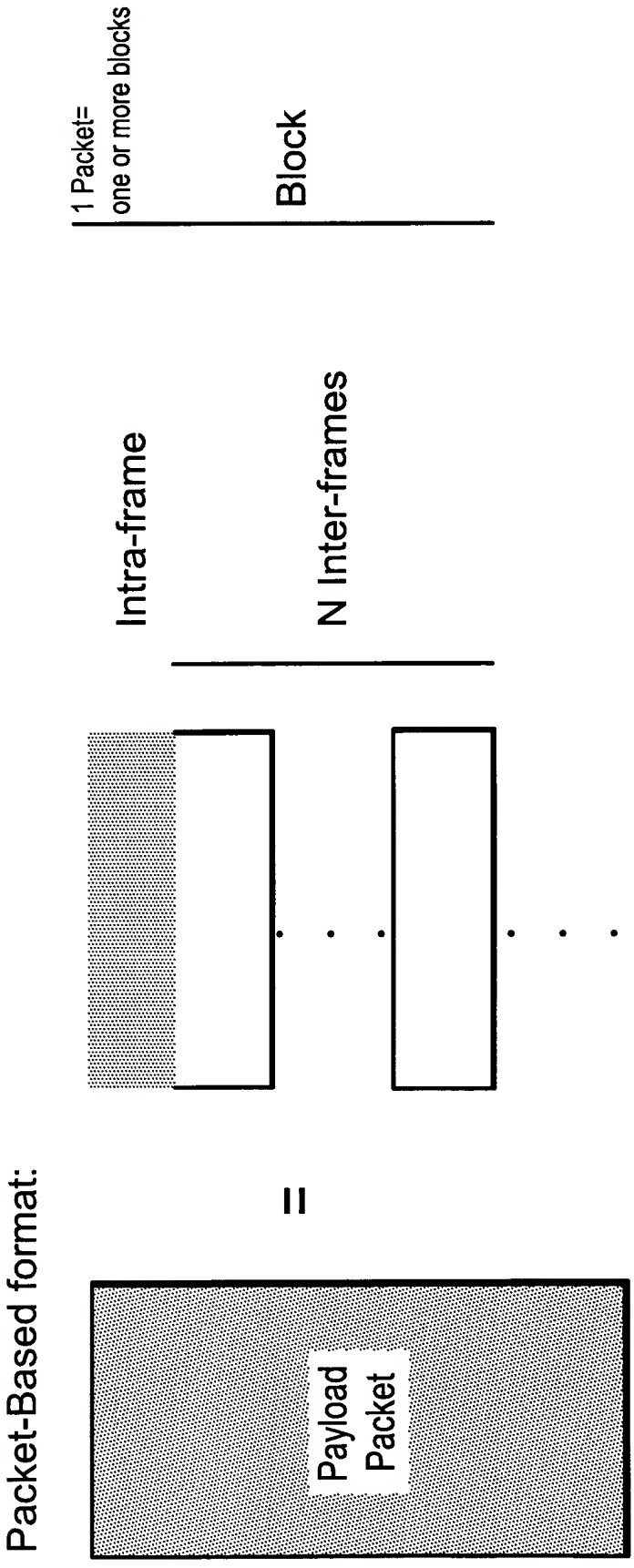
Packet-Based format:
=
Payload Packet
Intra-frame
N Inter-frames
Block
1 Packet= one or more blocks

FIG. 10

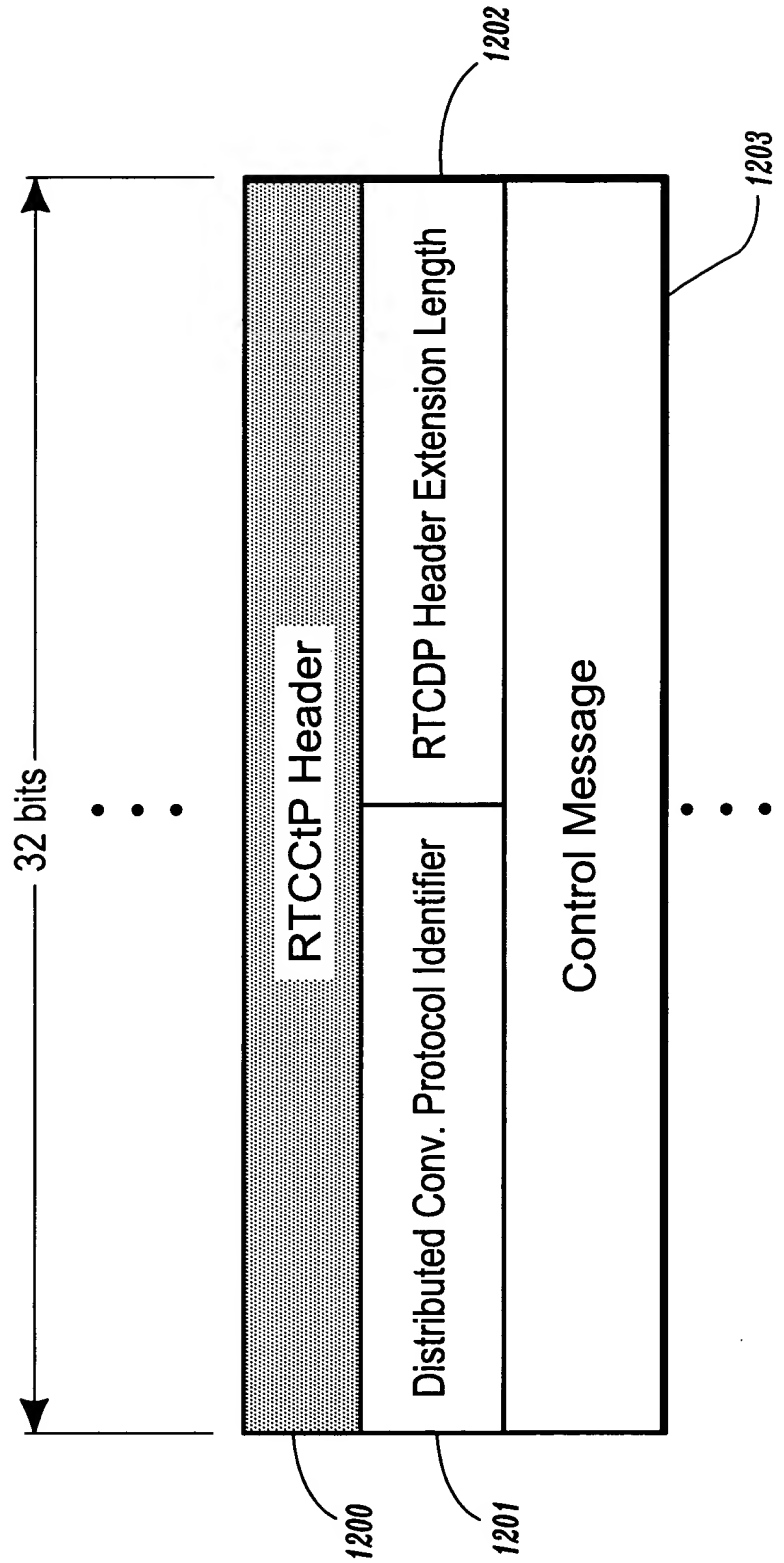
32 bits
RTCCtP Header
RTCDP Header Extension Length
Distributed Conv. Protocol Identifier
Control Message
1200
1201
1202
1203

FIG. 12

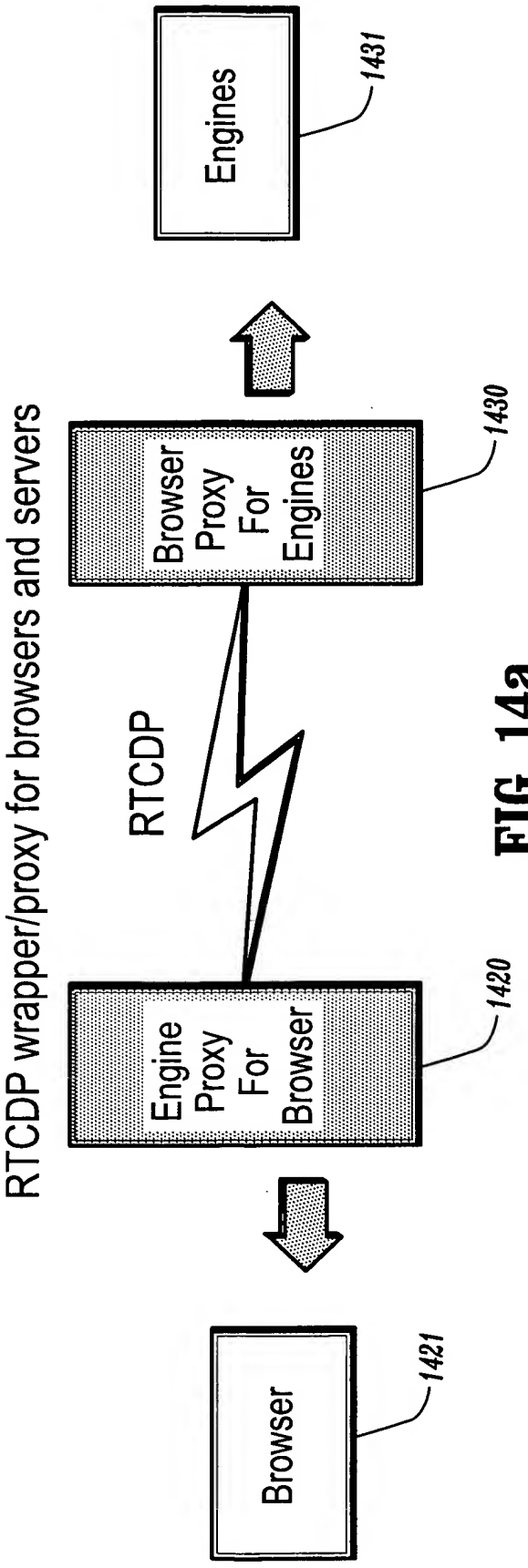

FIG. 14a

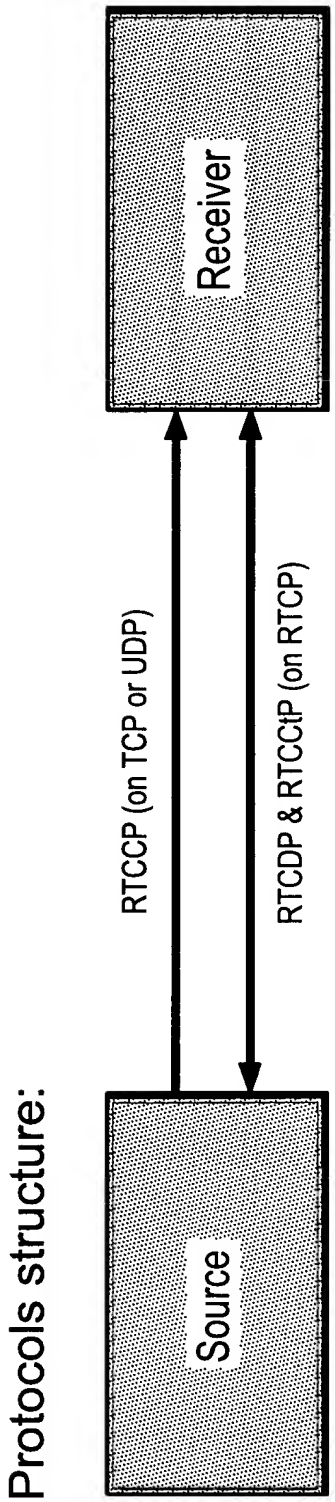

FIG. 13

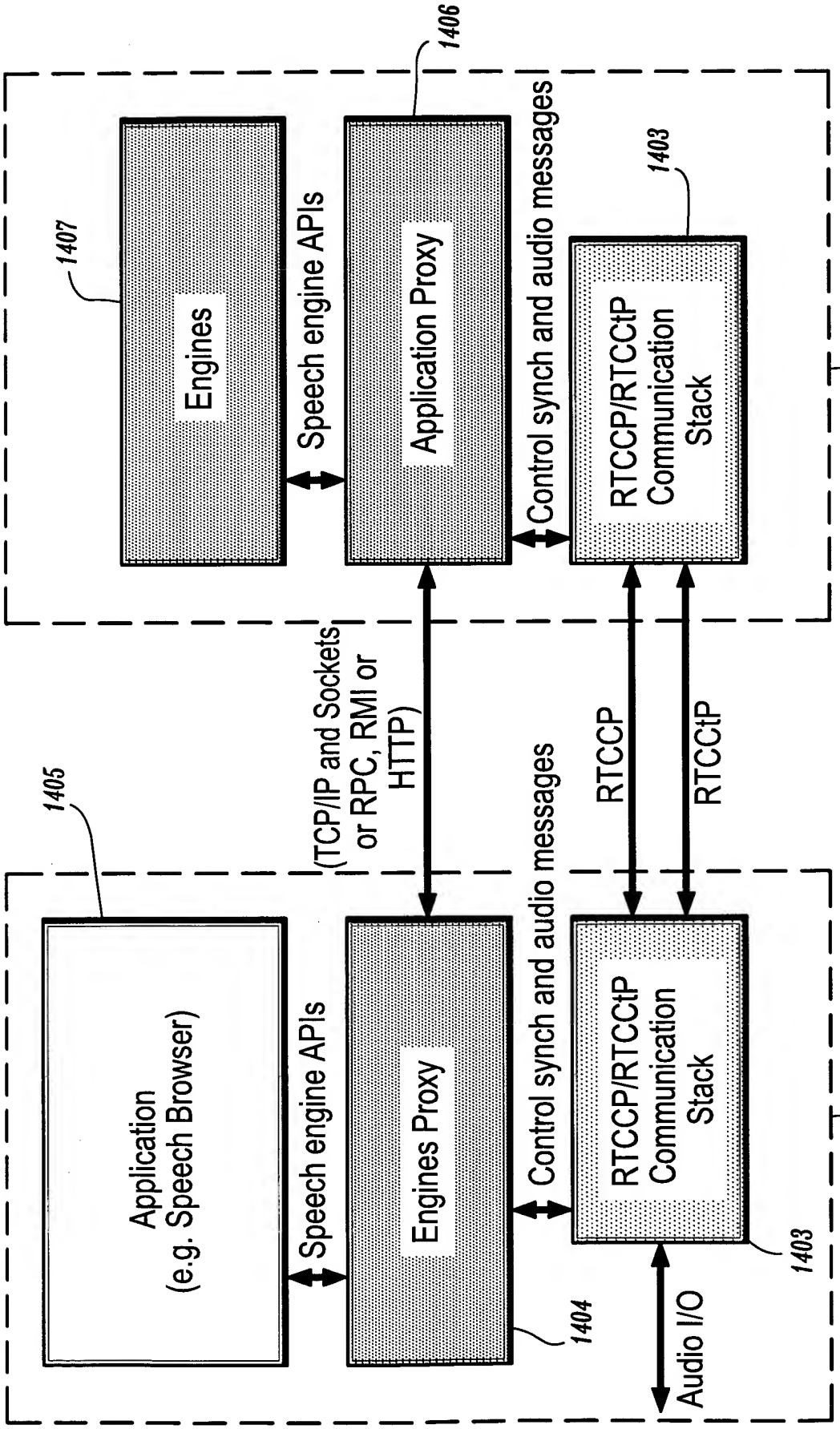
1406
1407
Engines
Speech engine APIs
Application Proxy
Control synch and audio messages
1403
RTCCP/RTCCtP Communication Stack
1402
(TCP/IP and Sockets or RPC, RMI or HTTP)
RTCCP
RTCCtP
1405
Application (e.g. Speech Browser)
Speech engine APIs
Engines Proxy
Control synch and audio messages
RTCCP/RTCCtP Communication Stack
1403
1404
Audio I/O
1401
1400

FIG. 14b

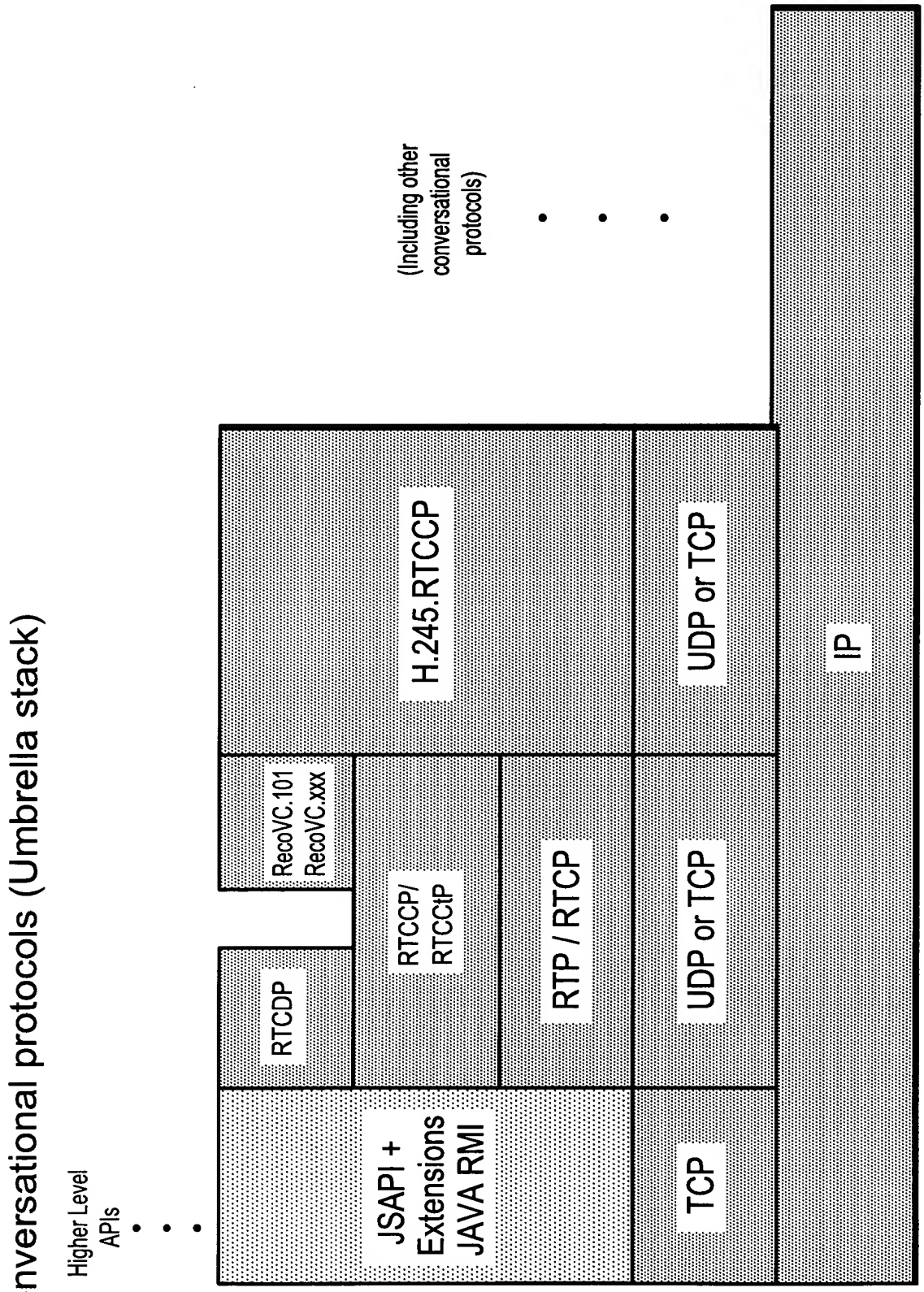
Conversational protocols (Umbrella stack)
Higher Level APIs
JSAPI + Extensions JAVA RMI
RTCDP
RecoVC.101
RecoVC.xxx
RTCCP/ RTCCtP
RTP / RTCP
H.245.RTCCP
(Including other conversational protocols)
TCP
UDP or TCP
UDP or TCP
IP

FIG. 15

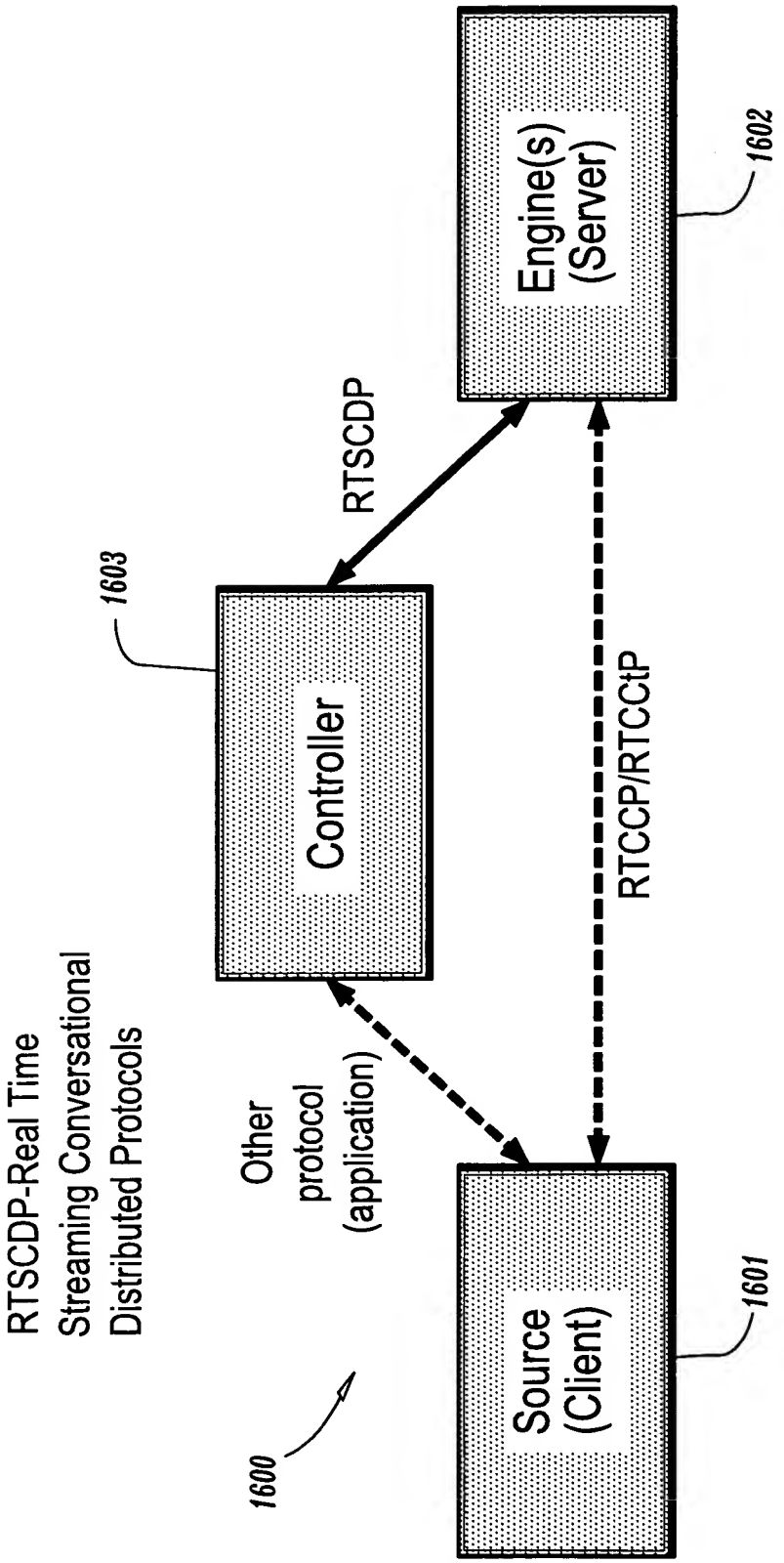
RTSCDP-Real Time
Streaming Conversational
Distributed Protocols
1600
Other
protocol
(application)
Source
(Client)
1601
Controller
1603
RTSCDP
RTCCP/RTCCtP
Engine(s)
(Server)
1602

FIG. 16

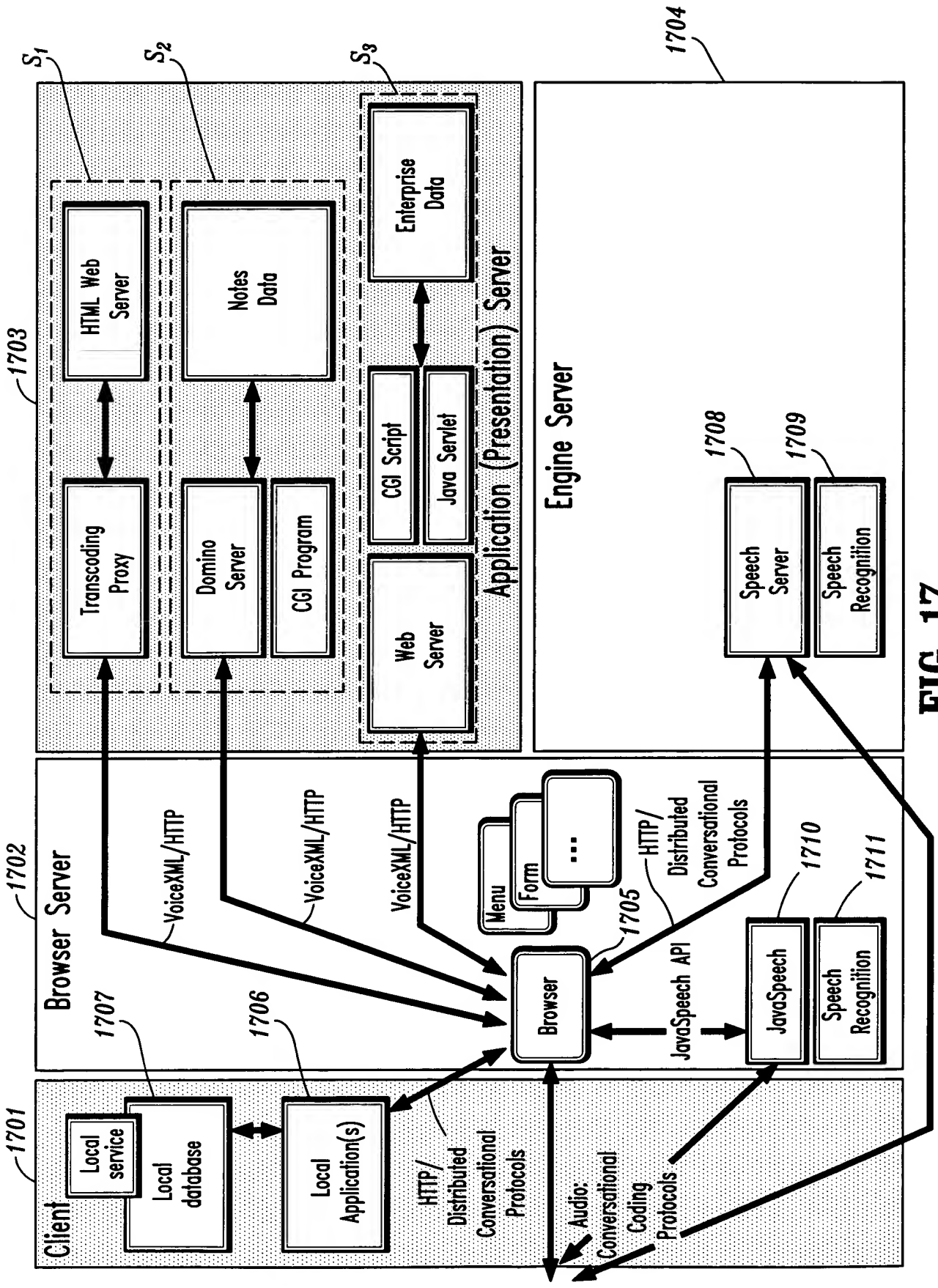
1701
Client
Local service
Local database
Local Application(s)
HTTP/ Distributed Conversational Protocols
Audio: Conversational Coding Protocols
1702
Browser Server
1707
1706
VoiceXML/HTTP
VoiceXML/HTTP
VoiceXML/HTTP
Browser
Menu
Form
...
1705
HTTP/ Distributed Conversational Protocols
JavaSpeech API
1710
1711
JavaSpeech
Speech Recognition
1703
Application (Presentation) Server
S_1
S_2
S_3
HTML Web Server
Notes Data
Enterprise Data
Transcoding Proxy
Domino Server
CGI Program
CGI Script
Java Servlet
Web Server
1704
Engine Server
1708
1709
Speech Server
Speech Recognition

FIG. 17

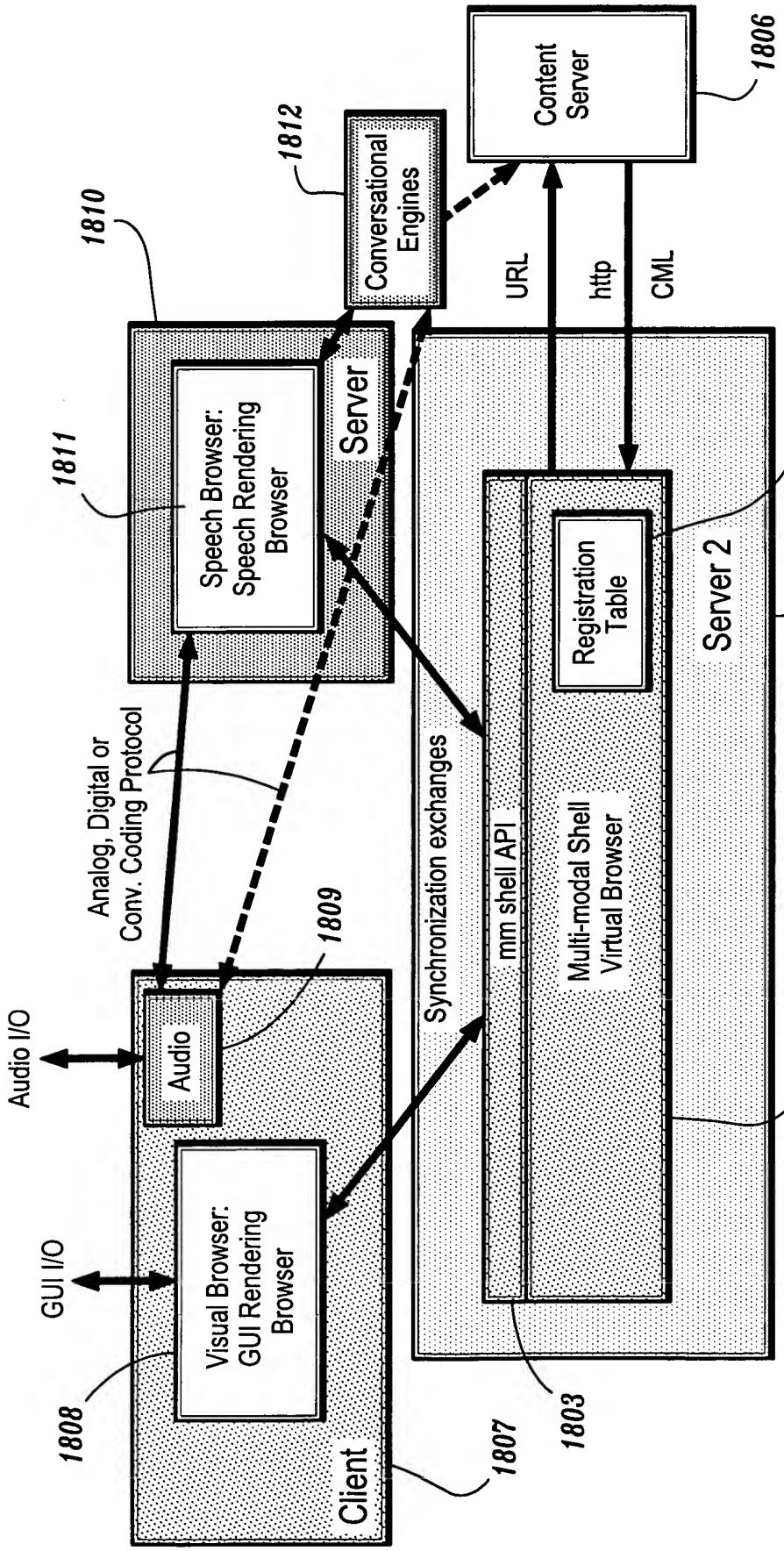
Content Server
1806
1812
Conversational Engines
1810
URL
http
CML
Server
1811
Speech Browser: Speech Rendering Browser
1804
Registration Table
Server 2
1805
Analog, Digital or Conv. Coding Protocol
Synchronization exchanges
mm shell API
Multi-modal Shell Virtual Browser
1809
Audio I/O
Audio
1801
GUI I/O
Visual Browser: GUI Rendering Browser
1808
Client
1807
1803
1800

FIG. 18

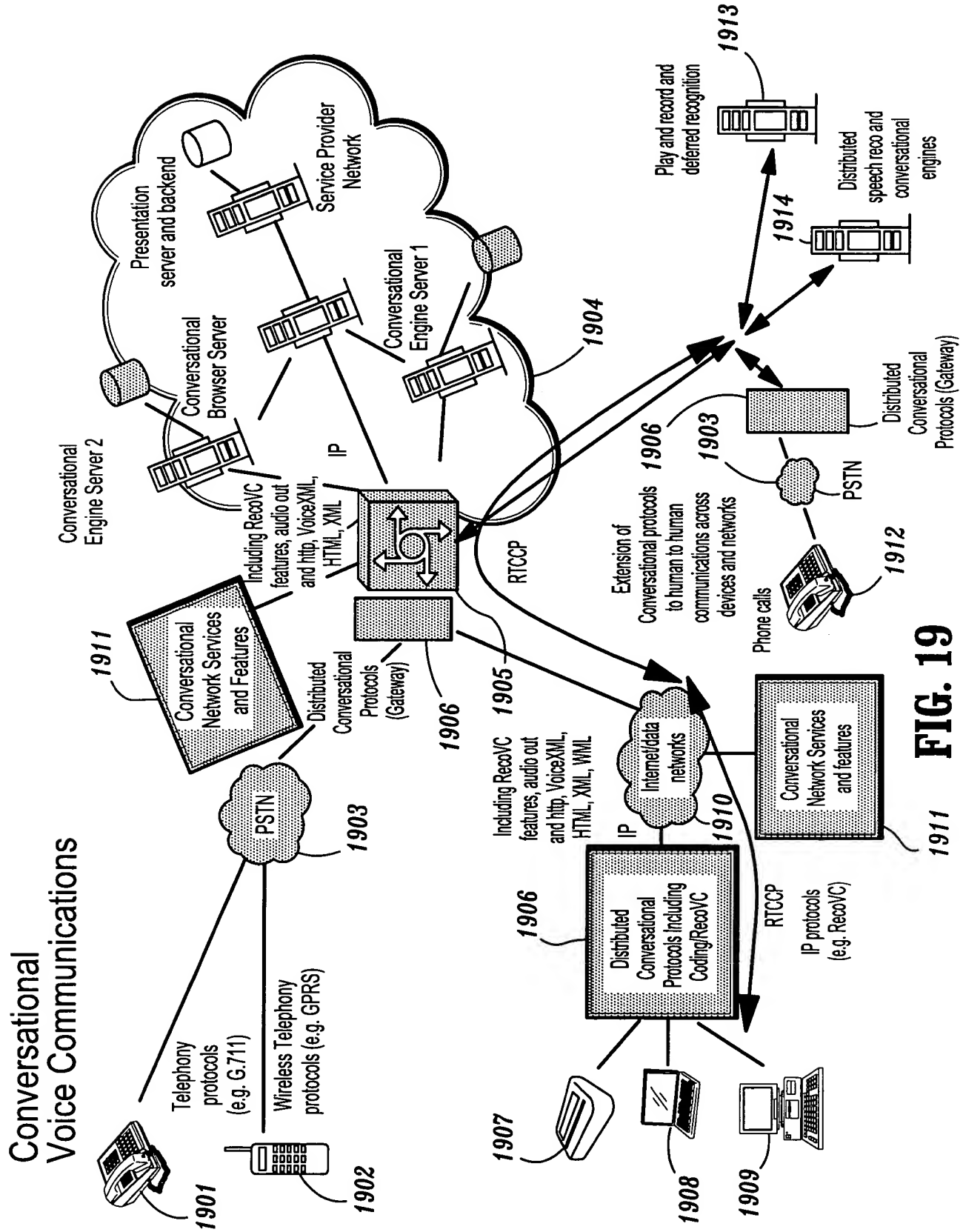
Conversational
Voice Communications
Presentation server and backend
Service Provider Network
Conversational Browser Server
Conversational Engine Server 1
Conversational Engine Server 2
1904
IP
Including RecoVC features, audio out and http, VoiceXML, HTML, XML
RTCCP
Play and record and deferred recognition
1913
Distributed speech reco and conversational engines
1914
Distributed Conversational Protocols (Gateway)
1906
1903
PSTN
Extension of Conversational protocols to human to human communications across devices and networks
Phone calls
1912
Conversational Network Services and Features
1911
Distributed Conversational Protocols (Gateway)
1906
1905
PSTN
1903
Including RecoVC features, audio out and http, VoiceXML, HTML, XML, WML
IP
Internet/data networks
1910
Conversational Network Services and features
1911
1906
Distributed Conversational Protocols Including Coding/RecoVC
RTCCP
IP protocols (e.g. RecoVC)
Telephony protocols (e.g. G.711)
Wireless Telephony protocols (e.g. GPRS)
1901
1902
1907
1908
1909

FIG. 19